

MCIEA School Quality Measures Reliability Analyses

Results from the 2016-2017 Student and Teacher Surveys

by Edward J. Kim

For its School Quality Measures (SQM) work, the Massachusetts Consortium for Innovative Education Assessment (MCIEA) uses multiple measures to track student and school performance beyond typical statewide evaluations in math and English. Specifically, the SQM draws on reports from students and teachers about the social and material conditions of their schools, as well as their own non-observable cognitive and emotional experiences. The content of the MCIEA perception surveys, administered to all teachers in the consortium, as well as to students in grades 4 and above, is aligned with the MCIEA School Quality Framework.

Despite broad public and scholarly consensus about the importance of the constructs included in the School Quality Framework, most of the measures used in the SQM are relatively new. This fact does not make them any less effective or valuable; it does, however, require careful attention to issues of validity and reliability. To that extent, this report presents an initial investigation into the MCIEA survey scales and items based on recent data compiled from the 2016-2017 administration of the survey.

A standard way to assess the properties of a survey scale is to measure its reliability.¹ Cronbach's α , as one such measure, represents the reliability across items, quantifying the degree to which each question in a scale seems to align with the other questions. Essentially, a high Cronbach's α suggests that the constituent items on a survey scale coherently capture the same construct. Though benchmarks for a sufficiently high Cronbach's α depend on the context and the intended interpretation, Nunnally (1978) suggests that a benchmark of 0.7 suffices for initial research purposes, such as identifying potential areas of concern, as is the purpose of this report.²

As shown in Table 1 and Table 2, the majority of MCIEA's survey scales demonstrate acceptable levels of reliability:

Table 1: Student Scales			
Scale name and SQM label	Cronbach's α	> α if item excluded?	Number of Items ³
1A.ii. Effective practices (4 th - 5 th) ⁴	0.8588	No	7
1A.ii. Effective practices (6 th - 12 th)	0.9257	No	7
2A.i. Student physical safety	0.6805	Yes: 0.6877	4
2A.ii. Student emotional safety	0.6236	No	3
2B.i. Student sense of belonging	0.8324	No	6
2B.ii. Student-teacher relationships (4 th - 5 th)	0.7667	No	5
2B.ii. Student-teacher relationships (6 th - 12 th)	0.8653	No	5

¹ American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

² Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

³ Note that by construction, Cronbach's α increases with more items; scales with fewer items will naturally tend towards lower reliability as captured by α despite have high internal coherence. Specifically, the calculation for Cronbach's alpha is $N \cdot r / (1 + (N-1) \cdot r)$, where r is the average inter-item correlation and N is the number of items.

⁴ Selected scales related to students' perceptions of their teacher varied by grade level. Students in grades 6-12, where subjects are taught by different teachers, were randomly given a specific teacher when responding to questions (math teacher, English teacher, science teacher, or social studies teacher).

Table 1: Student Scales			
Scale name and SQM label	Cronbach's α	> α if item excluded?	Number of Items ³
2C.i. Valuing of learning	0.8668	No	6
2C.ii. Academic challenge (4 th - 5 th)	0.6970	Yes: 0.7128	5
2C.ii. Academic challenge (6 th - 12 th)	0.8146	Yes: 0.8220	5
3A.ii. Content specialists and support staff	0.6696	NA ⁵	2
4B.i. Engagement in school	0.7514	Yes: 0.8319	3
5A.i. Appreciation for diversity	0.8287	No	5
5A.ii. Civic participation	0.7934	No	4
5B.i. Perseverance and determination	0.7664	No	5
5B.ii. Growth mindset	0.3744	Yes: 0.5978	3
5C.ii. Valuing creative and performing arts	0.6609	No	3
5D.i. Social and emotional health	0.7391	No	4

Table 2: Teacher Scales			
Scale name and SQM label	Cronbach's α	> α if item excluded?	Number of Items
1A.i. Professional qualifications	0.7314	No	3
1B.i. Professional community	0.8401	Yes: 0.8469	8
1B.ii.a ⁶ . Quality professional development	0.8873	No	4
1B.ii.b. Teacher collaboration	0.7670	No	3
1B.iii.a. Teacher-principal trust	0.9191	No	4
1B.iii.b. Instructional leadership	0.9316	No	4
2A.ii. Student emotional safety	0.8319	Yes: 0.8771	3
2B.i. Student sense of belonging	0.8925	No	4
3A.i. Physical space and materials	0.7879	No	5
3A.ii. Content specialists and support staff	0.8257	No	4
3B.i. Curricular strength and variety	0.7829	No	4
3B.ii. Cultural responsiveness	0.8606	No	4
3C.i. Family-school relationships	0.7977	Yes: 0.8697	4
3C.ii. Community involvement	0.8602	No	4
4A.ii. Overall student performance	0.7683	Yes: 0.8827	3
4C.i. Problem-solving emphasis	0.7643	No	3
5C.i. Participation in creative and performing arts	0.8412	Yes: 0.8439	3
5D.ii. Physical health	0.7903	Yes: 0.7988	3

Of the 36 individual scales, we find that 30 have an overall Cronbach's α higher than 0.7, though most of the remaining are quite near this benchmark. Eleven scales contain an item that, if excluded, would increase the scale's Cronbach's α , implying that the item is notably dissimilar to other items in the scale. This may suggest that the item should be removed, especially if the potential difference is large (on the student survey, for example, Cronbach's α for the "Growth Mindset" (5B.ii) scale would rise from 0.3744 to 0.5978 by dropping the lowest-performing item). However, if an item's inclusion were critical for accurately capturing a construct, this may justify retention despite its relative dissimilarity.

⁵ Cronbach's alpha if an item is excluded is uninterpretable with only two items composing the scales

⁶ The measures for "Support for growth and development" (1B.ii) and "Effective leadership" (1B.iii) were comprised of two unique scales. The reliability for each scale is included here.

The results of this report suggest that a substantial majority of the scales included in MCIEA's student and teacher perception surveys have been captured reasonably well. Only one scale arose as particularly problematic: "Growth Mindset" (5B.ii). Additional work, beyond measuring reliability as measured by Cronbach's α , will be necessary to ultimately justify the valid application of any of these scales.

Partly in response to these results, the SQM team has made minor changes for the 2017-2018 school year administration of the survey. Specifically, the wording of the questions in the scale "Growth Mindset" (5B.ii) was revised for clarity, and a new question was added to better represent the construct. The team similarly adjusted the wording of questions in other scales to improve reliability. These included the scales for "Student Emotional Safety" (2A.ii), "Engagement in School" (4B.i), and "Appreciation for Creative and Performing Arts" (5C.ii). In addition, in response to stakeholder feedback, the team added items to the student survey related to "Cultural Relevance" (3B.ii) and "Academic Stress" (5D.i)⁷, as well as items for the teacher survey related to "Academic Challenge" (2C.ii).

The MCIEA SQM project addresses an important gap in current educational practice and literature. Schools occupy a considerable role in adolescent development beyond what is currently measured and reported. Without intentional efforts to measure these additional impacts, we ignore an opportunity to better understand the complex relationship between students and schools, and possibly improve the efficacy of the school system. This report corroborates previous analyses⁸ that find the surveyed constructs in MCIEA's SQM work demonstrate acceptable levels of internal coherence. Continued work on this project will likely provide important data on educational practices in Massachusetts schools, and valuable insight into the topic of noncognitive impacts and outcomes.

⁷ Note the numbering and wording of these scales differs slightly from the 2016-2017 version of the survey

⁸ Kelly, M.P., Feistman, R., Schneider, J., & Noonan, J. (2017). Student Survey-Based Measures of School Quality. Boston: Center for Collaborative Education. Retrieved from http://www.mciea.org/images/PDF/Student_Survey-Based_Measures_of_School_Quality.pdf; Gagnon, D. J., & Schneider, J. (2017). Holistic school quality measurement and the future of accountability: Pilot-test results. *Educational Policy*, 1–27. <https://doi.org/10.1177/0895904817736631>